Crop Yield Prediction using Data Mining Approach

Sushmita Narayan Gaonkar[1], V. Deepika[2]

[1]*Student, Department of CSE, NHCE Bangalore, India*
[2]*Sr. Assistant Professor, Department of CSE, NHCE Bangalore, India*

Abstract: **Prediction of yield is immensely popular among farmers as this knowledge contributes to appropriate selection of crops for sowing. It serves as a crucial challenge since it requires meticulous study of huge data, thus making a perfect candidate for application of data mining. This paper presents the different techniques in data mining that can be used for predicting the features of environment dataset. The prediction algorithm that will be compared are Naive Bayes and K-Nearest Neighbor, Decision Tree, fuzzy logic and Neural Network**

Keywords: **Yield Prediction, Data Mining, Naive Bayes, K-Nearest Neighbor, Decision Tree, fuzzy logic and Neural Network.**

1. Introduction

In recent years, data mining has a huge impact on information technology and has become an important aspect of our life. In the agricultural field, data mining is considered as a novel research field. "Data Mining is a process of discovering previously unknown and likely impressive pattern in large dataset". The main task of data mining is to inspect and analyze a large amount of data and classify the data into clusters or groups (predefined groups). Data Mining is used to explore huge data sets and launch helpful classifications and patterns in data sets. The aim of the data mining procedure is to mine the information data from a data collection and change it into an explainable framework for unknown past, and a potentially growing model in huge database. Data mining is generally grouped as predictive and descriptive type. But in farming areas, predictive type is typically used. Data mining technique has been divided in two groups:

- Classification
- Clustering
- *Classification:* It is the process of placing or assigning the categorical variables into predefined classes. An algorithm needs to be selected to place these data in categories. The decision rules are based on training data and then used to locate these data in pre-determined groups. The rules are further validated by the validation dataset.
- *Clustering:* The data items are clustered according to their logical relationships or natural groupings and a structure as a whole is generated. There are no pre-defined groups, thus, clustering comes in the group of

undirected Data Mining techniques. Each cluster is a collection of homogeneous elements, which may be exclusive to that group, but are similar to each other. K-means clustering is a simple clustering method that has been used in similar research.
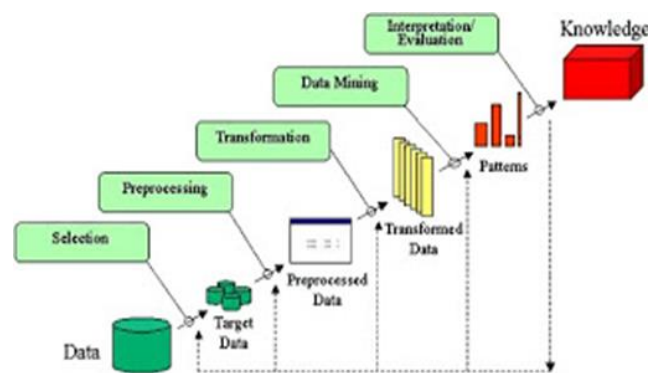

Fig. 1. KDD process in data mining

Data mining procedure is separated into seven methods:
- Data cleaning
- Data integration
- Data selection
- Data transformation
- Data mining
- Pattern estimation
- Knowledge display

2. Methodology

Crop prediction system has following fields:

A. Crop database

Dataset consist of information regarding agriculture field, soil (Phosphorus (P), Nitrogen (N), Potassium (K) content and micronutrients), moisture, rainfall and humidity in selected area.

B. Input data

Data related to agriculture field which has been stored in database such as humidity, rainfall, moisture, information related to soil values of such various variables, has been given as input.
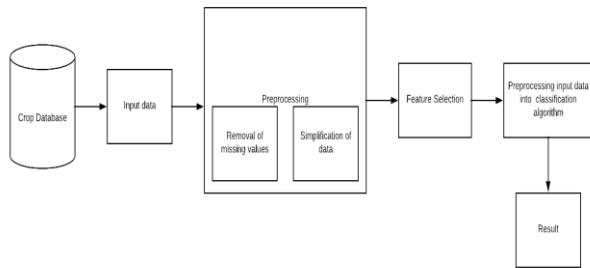
International Journal of Research in Engineering, Science and Management
Volume-2, Issue-1, January-2019
www.ijresm.com | ISSN (Online): 2581-5792

158

Fig. 2. Crop prediction system

C. Preprocessing

Since data is acquired from various resources, data might be raw information or might have missing values or redundant data or may be incomplete data, hence data has to be preprocessed.

D. Feature selection

The main aim of feature selection is to identify the most relevant features from dataset, so that irrelevant data has been removed.

E. Classification algorithm

Efficient and appropriate algorithm should be used for classification. Algorithm such as naïve Bayes can be used.

F. Result

Recommendation or prediction can be provided to farmers based on obtained results.

3. Conclusion

Our study specifies how improvements can be done in the fields of agriculture information. To carry out this study, we have used classification algorithm to gain the accuracy of prediction. We conclude that Naïve Bayes is most efficient for prediction of the yield of the crops.

4. Future Scope

The future scope is to create more efficient and appropriate models using classification methods like support vector machine (SVM). This paper uses a dataset that is comparatively smaller in size due to some system complexities. Further a huge dataset could be used to perform these tasks.

References

[1] Yogesh Gandge and Sandhya, "A Study on Various Data Mining Techniques for Crop Yield Prediction," in ICEECCOT, Mysore, 2017, pp. 420-423.
[2] Monali Paul, Santoshn K and Ashok Verma, "Analysis of Soil and predition of Crop Yield using Data Mining Approach," in ICCICN, Jabalpur, 2015, pp. 766-711.
[3] R. Sujatha and P. Isakki, "A Study of Crop Yield Forecasting Using Classification Techniques," in ICCTIDE, Kovilpatti, 2016.